# A case study of excessive subtropical transport in the stratosphere of a data assimilation system

Wei Wu Tan,[1,2] Marvin A. Geller,[3] Steven Pawson,[4,2] and Arlindo da Silva[2]

Received 8 August 2003; revised 18 March 2004; accepted 7 April 2004; published 2 June 2004.

[1] Assessments of transport show that results derived from assimilated (or analyzed) winds exhibit significantly larger mixing and entrainment rates compared to those derived from general circulation model (GCM) winds, where the GCM transport is somewhat closer (statistically) to values inferred from observations. This discrepancy presents a challenge to our ability to understand and model the global distributions of long-lived trace gases. We use the Goddard Earth Observation System's Finite Volume Data Assimilation System to explore this issue by examining how the data assimilation process alters the dynamics of the underlying GCM and how this leads to greater lower stratospheric mixing and transport in the subtropics. We show that the excessive subtropical transport is related to the proliferation of eddy features in the subtropics, and we examine various possibilities that may cause this. These include the generation of upward propagating features, equatorward propagating features, and meridionally confined features in the process of data assimilation. In particular, it is argued that unstable regions forced directly by the analysis increments play an important role in generating the excess subtropical transport. *INDEX TERMS:* 0341 Atmospheric Composition and Structure: Middle atmosphere—constituent transport and chemistry (3334); 3334 Meteorology and Atmospheric Dynamics: Middle atmosphere dynamics (0341, 0342); 3337 Meteorology and Atmospheric Dynamics: Numerical modeling and data assimilation; *KEYWORDS:* subtropical transport barrier, data assimilation, stratosphere

**Citation:** Tan, W. W., M. A. Geller, S. Pawson, and A. da Silva (2004), A case study of excessive subtropical transport in the stratosphere of a data assimilation system, *J. Geophys. Res.*, *109*, D11102, doi:10.1029/2003JD004057.

## 1. Introduction

[2] The study of transport barriers in the stratosphere is recognized as one of the most important issues in our understanding of global changes. It is now well established that there are two such barriers of transport in the winter lower stratosphere. They correspond to the polar vortex edge and the subtropical edge of the midlatitude surf zone. The importance of the polar transport barrier is very well recognized, as it is a necessary condition for the formation of the springtime Antarctic ozone hole [e.g., *Schoeberl and Hartmann*, 1991]. On one hand, the subtropical transport barrier prevents anthropogenic trace gases rising from the troposphere in the tropics from mixing effectively into the midlatitude lower stratosphere. On the other hand, it limits the incursion of trace gases from midlatitude lower stratospheric sources into the tropics, where they will be brought quickly aloft and transported globally by the meridional circulation into the more photochemically active upper stratosphere.

[3] Numerous efforts have sought to understand the formation of these transport barriers, in other words, to understand the how and why of the formation and confinement of the surf zone [e.g., *Juckes and McIntyre*, 1987; *Juckes*, 1989; *Polvani et al.*, 1995]. In the context of the shallow water model, *Polvani et al.* [1995] argued that the formation of these transport barriers during the easterly phase of the quasi-biennial oscillation (QBO) is due to the interaction of the breaking Rossby waves with a thermal (radiative) relaxation to an equilibrium state. The breaking of the Rossby waves leads to the tightening of the potential vorticity (PV) contours in the polar region, with the size of the vortex regulated and maintained by the thermal drive, without which a realistic size of the vortex over time cannot be obtained. This idealized model shows that in the subtropics, thermal relaxation creates strong subtropical wind shear that prevents the breaking of the waves originating from the midlatitudes into the tropics.

[4] These transport barriers are now well documented by satellite observations of various trace gases and volcanic aerosols [e.g., *Trepte and Hitchman*, 1992; *Grant et al.*, 1996; *Mote et al.*, 1996, 1998; *Gray and Russell*, 1999; *Neu et al.*, 2003]. Moreover, the stratospheric QBO in the tropics also points to the confinement of the surf zone since, if

[1]Science Applications International Corporation, Beltsville, Maryland, USA.
[2]Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, Maryland, USA.
[3]Institute of Terrestrial and Planetary Atmospheres, State University of New York at Stony Brook, Stony Brook, New York, USA.
[4]Goddard Earth Science and Technology Center, University of Maryland Baltimore County, Baltimore, Maryland, USA.


0148-0227/04/2003JD004057

easterly waves from midlatitudes break freely into the tropics, the known morphology of the QBO, especially the westerly phase, will be altered.

[5] Since the effort to understand, assess, and reliably predict the impact of anthropogenic trace gases on the atmosphere culminates in the use of three-dimensional numerical models, a logical question to ask is whether these transport characteristics exist in various general circulation models (GCMs) and if they do, how well they resemble those in the real atmosphere. It should be emphasized that the correct representation of the transport characteristics, especially the transport barriers, as embodied in the wind fields in a GCM, plays an important role in our ability to model the three-dimensional evolution of trace gases in both the "online" approach in which the dynamical evolution is coupled through radiative feedback to the chemistry, and the "offline" approach in which the wind field is used merely as an advective agent. Since both of the main elements in the formation of the transport barriers, namely the potential vorticity restoration mechanism and thermal (radiative) relaxation, are present in GCMs, such models contain the necessary physical processes needed to capture some aspects of the barriers.

[6] Our ability to simulate the atmosphere using GCMs is limited by many factors. Among the most prominent are the following: the chaotic nature of the atmospheric motions, which causes medium-range forecasts to "drift" from the true atmospheric state, and imperfections in our ability to represent physical processes through parameterizations (which cause climate bias in free-running models). Both of these are impacted by limitations in computing resources, which restrict the resolution and the complexity of the parameterizations used.

[7] The limitations of pure model simulations, especially the drift from the true atmospheric state, which makes direct evaluation with observations almost impossible, support the use of data assimilation. Confrontation of the models with data offers two distinct benefits. First, the fields are constrained to resemble our best knowledge of the real world (with the inclusion of data and their error characteristics). Second, the model-data differences provide statistics for monitoring performance and, eventually, provide measures for improving our representation of processes in the models. Modern data assimilation techniques are based around GCM forecasts and statistical models of observation and forecast error covariances, which allow optimal coupling of forecasts and observations. During the assimilation process the GCM is essentially "forced" with a set of incremental fields that are statistically weighted differences between model prediction and observation. Since this process of data insertion episodically disturbs the balance of the dynamical equations governing the evolution of the GCM fields, significant differences in the characteristics of the dynamical fields might arise owing to the generation of noise by the imbalance. Also, the effect of unsatisfactory representation of the physical processes might be amplified during the assimilation [*Molod et al.*, 1996]. More importantly, all present methods of data insertion inevitably lead to the violation of basic fluid dynamical principles, notably the conservation of PV and mass. The absence of a balance condition and the sparsity of wind observations lead to particular challenges in data assimilation in the tropical stratosphere. It has been shown that subtropical transport varies substantially among different assimilated data sets [*Waugh*, 1996]. This arises from the scarcity of global wind observations, model biases within the tropics, and the methods of data assimilation. Therefore an investigation of the transport characteristics of a particular data assimilation system and comparison with the parent GCM, along with developing a quantitative understanding of the cause of the differences, is an important research task. In this sense a study of the subtropical transport barrier serves both as a purpose in itself and as a tool to test our ability in global modeling.

[8] Recent studies using effective diffusivity diagnostics calculated from models of global isentropic tracer advection driven by assimilated winds demonstrated that mixing and transport weaken substantially in the subtropics [e.g., *Haynes and Shuckburgh*, 2000a, 2000b; *Allen and Nakamura*, 2001]. This shows that some sort of barrier indeed exists in data assimilation systems. However, while effective diffusivity is useful in constructing the relative latitude dependence of mixing, it remains a challenge to relate it to actual tracer flux [*Haynes and Shuckburgh*, 2000b]. As a result these studies could not give a measure of how realistic these barriers are. *Rogers et al.* [1999] compared Improved Stratospheric and Mesospheric Sounder measurement of the Pinatubo aerosol with results from an offline three-dimensional transport model (SLIMCAT) driven by the UK Met Office (UKMO) and European Centre for Medium-Range Weather Forecasts (ECMWF) analyzed wind and temperature. They found that SLIMCAT's integrations, particularly runs driven by the UKMO analyses, produced too much transport of midlatitude air into the tropics. Other studies using trajectory-related methods as well as chemical transport models (CTMs) [e.g., *Douglass et al.*, 2003; *Schoeberl et al.*, 2003] show that assimilated winds drive significantly larger subtropical transport compared to winds from GCMs. In particular, Douglass et al. showed that compared to Finite Volume Data Assimilation System (FVDAS) winds, Finite Volume General Circulation Model (FVGCM) winds give results that are closer in their behavior to some aspects of those derived from observations. One particular problem pointed out in these studies is the rapid mixing of air from the midlatitudes into the tropics. A promise of CTMs driven by assimilated meteorological fields has been their anticipated ability to reproduce observed time series of atmospheric constituents (e.g., ozone) with known chemical and microphysical treatments, thereby giving confidence in the predictive ability of such models. Deficiencies in transport circulations from data assimilation compromise this promise.

[9] Nonetheless, there have been few, if any, studies that examine and compare a data assimilation system and its parent GCM to clarify the roles of data assimilation in the differences of lower stratospheric subtropical transport. This study attempts to take a step forward in that direction. We conduct a detailed case study of the Goddard Earth Observing System (GEOS) FVDAS and its parent GCM to provide a more concrete understanding of how data assimilation leads to greater subtropical transport in the lower stratosphere. In order for the dynamics fields in the data assim-

ilation system (DAS) and the GCM to be comparable, we focus on a timescale of ~1 month.

## 2. Models and Methodology

### 2.1. GCM and DAS

[10] The GCM used in this study is based on the flux form semi-Lagrangian transport scheme [*Lin and Rood*, 1996, 1997], formulated with finite volume elements with a quasi-Lagrangian vertical coordinate [*Lin*, 1997; S.-J. Lin, A "vertically Lagrangian" finite-volume dynamical core for global models, submitted to *Monthly Weather Review*, 2003]. This eliminates the need for explicit calculation of vertical advection, since the motion of the material surfaces is implicitly described in this framework. The numerical construction [*Lin and Rood*, 1996] ensures an accurate representation of transport by the resolved scale flow. The model is referred to as FVGCM. Physical parameterizations are from version 3 of the National Center for Atmospheric Research Community Climate Model [*Kiehl et al.*, 1998]. We used a configuration with 55 vertical levels extending from the surface to 0.01 hPa; while dynamical processes and transport are determined on the quasi-Lagrangian vertical surfaces, fields are remapped to hybrid sigma pressure surfaces for the computation of physical tendencies. The horizontal resolution is 1° in latitude and 1.25° in longitude.

[11] The data assimilation system used in this study is version 1.2r5 of the FVDAS, which is a prototype version of NASA's operational GEOS-4 DAS. It uses the Physical Space Statistical Analysis Scheme [*Cohn et al.*, 1998] and intermittent data insertion. Radiosondes and aircraft provide in situ data through the troposphere and into the stratosphere, while TIROS Operational Vertical Sounder level 1b radiances are used to constrain the global thermal structure, using the one-dimensional variational method of *Joiner and Rokke* [2000]. Analyses were performed with a 6-hourly cycle. The horizontal resolution of the analysis system used in this study is 2° in latitude and 2.5° in longitude. The data-model merger produces analyses of meteorological fields on the grid of the underlying model, which have been linearly interpolated to isentropic levels for diagnosis in this study. In particular, transport in the lower and middle stratosphere, on the 450-, 550-, 700-, 850-, and 1000-K theta surfaces, will be discussed. Between 30°S and 30°N these isentropes extend from 70 to 7 hPa (about 18–35 km).

[12] A series of experiments was performed for January and July in 1998, which are both in the easterly phase of the QBO. In January the analyses reveal easterly shear throughout the region of interest (Figure 1b). By July, westerly shear descended to the region above 20 hPa, and the equatorial easterlies are weaker. The discussion focuses on January, with comments on July only for comparative purposes.

[13] The corresponding model integrations are long-range forecasts. The January 1998 model run was initialized from FVDAS analyses on 31 December 1997 and was integrated for 1 month to cover the period of interest. This is clearly outside the range of predictability, so there will no longer be a unique correspondence between the model run and the analysis. However, for the tropical stratospheric winds the decay scale is much longer than that for (say) baroclinic disturbances [*Hamilton and Yuan*, 1992], so the January 1998 model run (see Figure 1) displays a closer resemblance to the true atmosphere than it would for a multiyear integration.

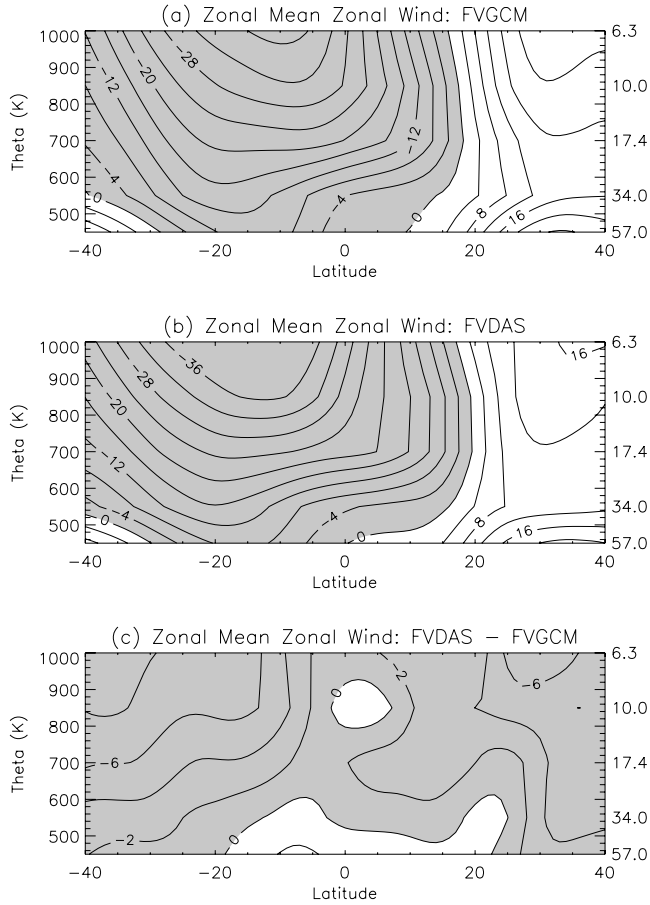


**Figure 1.** Zonal mean wind in January 1998, interpolated to isentropic levels for the (a) FVGCM and (b) FVDAS. The contour interval is 4 m $s^{-1}$, and easterlies are shaded. (c) Difference between FVGCM and FVDAS, which has a contour interval of 2 m $s^{-1}$. Shading denotes regions where the FVDAS winds are more easterly (or less westerly) than the model simulation. Pressure levels (in hPa) corresponding to the isentropes used in this study are shown on the right-hand axis.

### 2.2. Methods of Comparing Isentropic Subtropical Transport and Mixing

[14] To compare isentropic transport and mixing characteristics in the FVGCM and FVDAS, we use trajectory calculations as well as "equivalent length" [*Nakamura*, 1995, 1996] calculated from the PV fields. The main purpose is to quantify the differences in transport and mixing characteristics induced by the data assimilation process; we do not seek realistic estimates of quantities such as entrainment time, age of air, or diffusion coefficients. This justifies the use of a simple algorithm that quantifies transport from the midlatitudes into the tropics. Moreover, noise in the analyzed fields limits the usefulness of PV-based entrainment rate calculations using the differences in analyzed and advected PV values, such as the contour-crossing algorithm [*Sobel et al.*, 1997]. Unlike the

polar vortex, there is no subtropical minimum in metrics, based on PV, that one can use to identify the transport barrier [e.g., *Haynes and Shuckburgh*, 2000a; *Chen*, 1996]. In addition, the results in section 3.1 show that FVDAS equivalent length calculated from PV exhibits a double-peak structure in the subtropics. Such a double-peak structure appears also in the FVGCM; however, the equatorward peak is generally much smaller than that in the FVDAS. It will be shown that in the FVDAS the peak closer to the equator is related to eddy features generated in the process of data assimilation. This double-peak structure blurs the transport barrier associated with the large-scale dynamics, making it difficult to identify. Therefore, instead of trying to pinpoint the location of the transport barrier using metrics based on PV, we simply assume the transport barrier for each period of calculation to be covered by a latitude band close to the zero-wind line. This choice is corroborated by the results of *Neu et al.* [2003], in which they found that below 10 hPa the surf zone edge generally lies along the zero-wind line. In the summer hemisphere we simply used 20°S or 20°N as a reference latitude. Sensitivity studies with different reference latitudes will be discussed in section 3.1. The procedure adopted initializes a uniformly and globally distributed group of parcels on a $0.5^\circ \times 0.5^\circ$ grid and performs a series of 10-day back trajectory calculations [after *Schoeberl and Sparling*, 1995] using winds on the isentropes. We examine a 20° band located 10° equatorward of the reference latitude. The entrainment rate is defined as the percentage of parcels that started poleward of the reference latitude and ended in that band. The use of a 10° threshold ensures that parcels counted are indeed transported irreversibly. This threshold value is larger than the standard deviation of the equivalent latitude (calculated from PV) at the reference latitude. Using different threshold values (10°, 15°, and 20°), some extending 10° poleward of the zero-wind line, did not change the qualitative features of the comparison. When a larger threshold value was used, the rates decreased. This shows that a significant fraction of the parcels that moved into the tropics came from areas around the surf zone edge. With a longer time period, more parcels are able to travel a longer distance, and the entrainment rates increase correspondingly. Selected calculations with twice as many parcels show differences of <5% in the values of the entrainment rates. Even when the entrainment rates are originally very small (say, <1% per month), the differences induced by the change in resolution (or the number of parcels) were only as large as 8% of the rates. Within the framework of this particular algorithm our choice of the number of parcels is thus robust. Further discussion of sensitivity studies will be given when the results are presented.

[15] We supplement the entrainment rate diagnostics with equivalent length calculated from PV fields. This is done by the direct evaluation of the following equation [*Nakamura*, 1995, 1996; *Haynes and Shuckburgh*, 2000a, 2000b]:

$$L_e^2(\phi_e) = \langle |\nabla q|^2 \rangle / (\partial q/\partial A)^2 = L_0^2 \left[ a^2 \langle |\nabla q|^2 \rangle / (\partial q/\partial \phi_e)^2 \right], \quad (1)$$

where $\phi_e$ is the equivalent latitude derived from a tracer ($q$, in this case, the PV) with a monotonic latitudinal distribution, $L_0$ is the perimeter of the equivalent latitude at which $L_e$ is evaluated, $A$ is the area enclosed by the equivalent latitude, and $a$ is the radius of the Earth. In regions where stirring is small, the geometry of the tracer contours is simpler, and thus $L_e$ is closer to $L_0$, which is the lower bound of $L_e$. In equation (1), angle brackets denote the average around a given equivalent latitude. Although the values of $L_e$ depend on the grid resolution, the structure of extrema is independent of the resolution. As a result the latitudinal structure of $L_e$ is useful for identifying the strength and location of the stratospheric transport barrier [*Nakamura and Ma*, 1997]. Since our focus is in the subtropics, we will present the results as the ratio $\xi = L_e/L_0$ rather than $\log(\xi)$, as in global studies. A note of caution is that equivalent length is a rigorous measure of isentropic mixing only when it is calculated from a passive tracer advected in an isentropic transport model [e.g., *Haynes and Shuckburgh*, 2000a, 2000b; *Allen and Nakamura*, 2001]. In this paper, it was calculated using model/analyzed Ertel PV. To make sure that equivalent length calculated directly from the PV field itself indeed contains the qualitative features (a contrast of strongly mixed and weakly mixed regions) of isentropic mixing, we compare it with equivalent length calculated from isentropically advected PV fields obtained using the method of reverse domain filling (RDF). In this procedure, we use the 10-day backward trajectories mentioned in the preceding paragraph and record the model/analyzed PV fields at the end of the trajectories. These model/analyzed PV fields were truncated to zonal wave number 2 to remove fine-scale features. Since there was no diffusion during the advection, the RDF-constructed PV fields became very complex, and the equivalent length became very large, especially in the surf zone. Nevertheless, the contrast of regions with strong mixing and weak mixing is very pronounced. Moreover, qualitative differences of RDF equivalent length between the FVGCM and the FVDAS are similar to results calculated from model/analyzed PV directly. This affirms the qualitative validity of the equivalent length results presented in this paper.

## 3. Zonal Mean Wind, Entrainment Rates, and Equivalent Length

### 3.1. January 1998

[16] The simulated and assimilated zonal mean wind (Figure 1) reveal the same general structure with some subtle differences. In comparison to FVDAS wind the FVGCM wind exhibits easterly biases around the equator at the bottom of the region of interest and weak westerly biases at higher levels, apart from a small cell just north of the equator around 850 K. The zero-wind line is slightly closer to the equator at all levels above 550 K in the FVGCM than in the FVDAS. Since the structure of the zonal mean wind is highly relevant to the confinement of the surf zone [*Polvani et al.*, 1995], the similarity of the zonal mean wind in January between the two data sets makes this an appropriate case to study the influence of data assimilation on transport.

[17] The monthly mean analysis increment of the zonal mean wind calculated from the 6-hourly analyses exhibits large zonal anomalies. Figure 2 shows a representative scenario at 30 hPa. The average of the analysis increment around regions with wind data (sonde stations) is noticeably

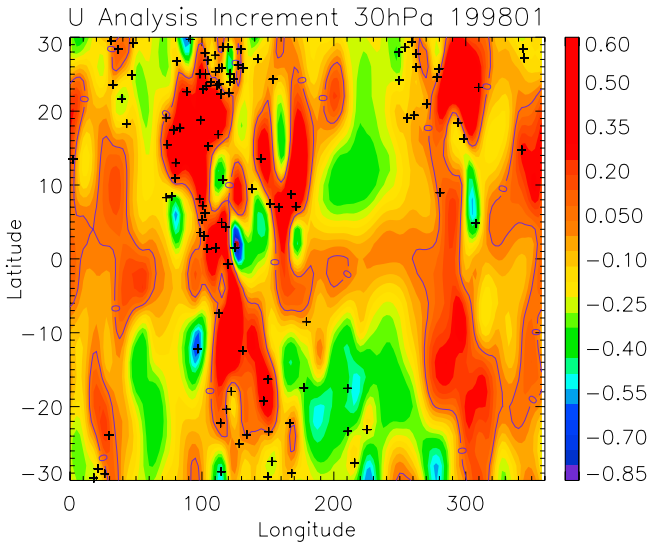


**Figure 2.** Monthly mean analysis increment of the zonal mean wind (m $s^{-1}$) calculated from the 6-hourly analyses in January 1998. Pluses mark locations of wind observations.

different from that in locations with no wind data. This is the result of model biases. One such known bias in the lower stratospheric tropics is the absence of the QBO in the FVGCM. As we will show in sections 4.2 and 4.4, the forcing of the model by the analysis increments plays an important role in generating a greater subtropical transport. One should note that this is not just a problem of using wind data. As long as the forcing of the model by analysis increments leads to an increase of wave power in the relevant regions, subtropical transport will be increased correspondingly. This will be elucidated in sections 4.2 and 4.4. The use of wind data is not the only cause, and experiments with wind data withheld are plagued with the same problem.

[18] Three 10-day backward trajectory calculations, initialized on 11, 21, and 31 January, were performed for each data set. Figure 3 shows entrainment rates obtained by summing the three 10-day calculations in January 1998, clearly showing that entrainment rates in the assimilated data set are much larger than those in the model (except at 1000 K), and both are dominated by the contribution from the winter (northern) hemisphere (except at 450 K). The vertical structure is also of interest. In the FVGCM the entrainment rates are the smallest at 450 and 550 K and do not exceed 1.1% at higher isentropes. In the FVDAS the entrainment rate is the largest at 550 K, decreases at 750 K, and increases again at 850 K. At 1000 K the rates in the two data sets are almost the same. This may be due to the fact that there is no direct wind observation above 10 hPa. The difference in altitudinal structure of the entrainment rates is essentially the same as the difference in equivalent lengths shown in the following paragraph. Note that the results in the FVGCM agree with various studies using trace gas observations which suggested a minimum in transport at around 500–550 K [e.g., *Mote et al.*, 1998; *Minschwaner et al.*, 1996]. This is not the case in the FVDAS. Plausible causes of the increased transport in the FVDAS will be discussed in section 4. In the FVGCM, there is no inflow of particles from the southern extratropics except at 450 K, and there is only a small amount of transport from the north, confirming the existence of robust subtropical barriers. In the FVDAS, entrainment rates from the north are much larger than those from the south, except at 450 and 1000 K. Note that entrainment rates in the summer hemisphere were calculated using a fixed reference latitude at 20°S. This contributes to the relatively large rates at 450 K. Sensitivity studies showed that at 450 K, entrainment rates in the summer hemisphere decreased when a reference latitude closer to the equator was used. Therefore the barrier in the summer hemisphere probably was very close to the equator at 450 K. However, regardless of the reference latitude used, summer hemisphere entrainment rates in the FVDAS are always larger than those in the FVGCM for this particular month. To account for the possibility that the actual transport barrier in the winter hemisphere may lie poleward of the zero-wind line, we extended the threshold poleward of the zero-wind line by 5° and 10° latitude, resulting in a threshold of 15° and 20°. In each case the rates decrease in both data sets, indicating that a significant amount of the material entrained into the tropics came from regions close to the edge. More importantly, these sensitivity studies showed that the differences between the two data sets are robust. We would like to point out that the results for each individual 10-day period show that, episodically, the entrainment rates in the FVGCM can be larger than those in the FVDAS.

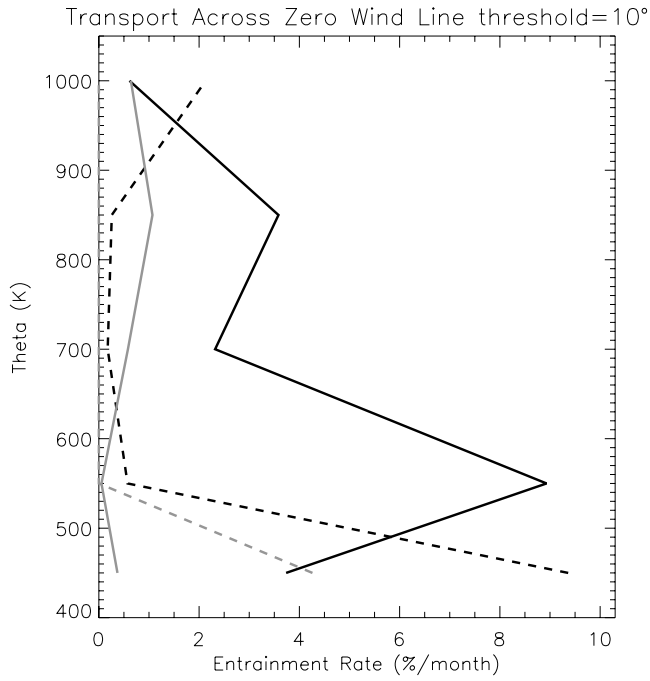


**Figure 3.** Entrainment rates obtained by summing three 10-day calculations in January 1998. Results for the FVDAS (black curves) and results for the FVGCM (grey curves) are shown. The rates from the winter (northern) hemisphere (solid curves) and summer (southern) hemisphere (dashed curves) are shown.

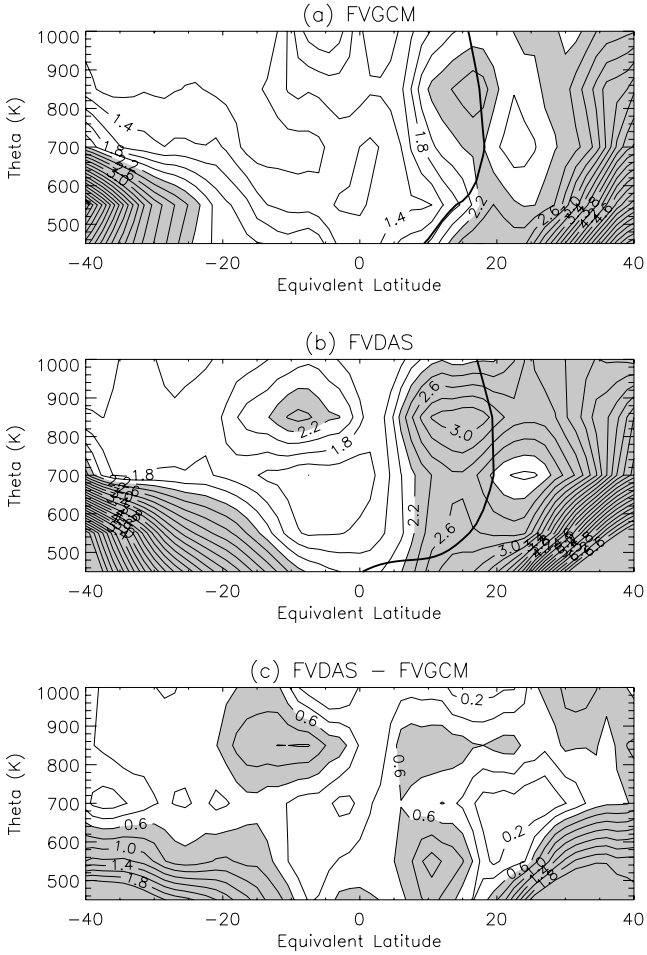


**Figure 4.** Monthly averaged equivalent latitude-potential temperature sections of $\xi$ for (a) FVGCM, (b) FVDAS, and (c) the difference between FVDAS and FVGCM. The contour interval in Figures 4a–4c is 0.2, and shaded region denotes values larger than 2.2 (Figures 4a and 4b) or differences larger than 0.6 (Figure 4c). Figures 4a and 4b show the zero zonal mean wind (thick, solid curves). Note that the values are always larger in the FVDAS than in FVGCM.

This is the case for the calculation initialized on 31 January at 850 K, where the rate in the FVGCM is 1.07%, and that in the FVDAS is 0.56%. Except for rare episodes like this, entrainment rates in the FVDAS are consistently larger than those in the FVGCM. While only 1 month of calculations is presented here owing to the lack of a data set covering a longer time span, the result of greater transport and mixing in the FVDAS is supported by other studies using longer data sets [e.g., *Tan*, 2000; *Douglass et al.*, 2003; *Schoeberl et al.*, 2003].

[19] Figure 4 shows the monthly average of $\xi$ for the FVGCM and FVDAS as well as their difference. The abscissa is equivalent latitude. Thick, solid curves show the zero-wind line. We see that in Figure 4a, $\xi$ has a band of minima north of 20°N in equivalent latitude, extending throughout the altitudes of interest. This band of minima is associated with a band of strong PV gradient poleward of the zero-wind line. While the band of strong PV gradient is often regarded to be the location of the transport barrier, there is no way to ascertain that this band of minima indeed corresponds to the actual transport barrier. Regardless, this band is rather permeable in this case since southward propagating stationary planetary waves can episodically break across it when encountering the zero-wind line equatorward. In contrast, note that in the confined regime of the idealized shallow water model experiments of *Polvani et al.* [1995] the band of strong PV gradient related to the strong subtropical wind shear is located equatorward of the zero-wind line. In that idealized case the subtropical transport barrier is impermeable since southward propagating stationary planetary waves break around the zero-wind line poleward of the steep PV gradient without penetrating it. In Figure 4a, stirring occurs around the band of maxima south of 20°N, indicated by the narrow shaded region. In contrast to the Northern Hemisphere, equivalent length in the Southern Hemisphere is relatively small and devoid of structures. This is the result of the absence of planetary wave activity in the summer hemisphere. In Figure 4b, for the FVDAS, the structure of minima and maxima are similar to those in the FVGCM but with different shapes and locations. In particular, we would like to point out that although the zero-wind line in the FVDAS is located further to the north compared to that in the FVGCM, the band of maxima (shaded region centered at 15°N) in the subtropics is located to the south of the corresponding feature in the FVGCM. This indicates that there are probably some extra processes absent in the FVGCM that are responsible for these features. As we will argue in section 4, the appearance of these maxima in the FVDAS is related to data insertion. Equivalent length calculated from isentropic advection of zonally truncated PV contours using the RDF technique indicates that these peaks will likely exist even in an isentropic transport model of a passive tracer. We also note that equivalent length at 600 and 800 K of the same period calculated by *Ma et al.* [2003] using $N_2O$ and $CH_4$ from CTMs driven by GEOS (operational system of FVDAS) analyzed wind showed similar noise structure in the subtropics. Effective diffusivity calculated by *Haynes and Shuckburgh* [2000a] using the ECMWF analysis also showed some extension of higher diffusivity values into the subtropics, although not as pronounced as that from *Ma et al.* [2003] and in our case. It is not clear if these structures in different calculations are driven by the same cause without further investigation. In the southern subtropics a cell of relatively large equivalent length centered around 8°S appears in the upper half of Figure 4b. The difference of $\xi$ between FVDAS and FVGCM (Figure 4c) is positive everywhere, indicating that the assimilation process has resulted in more noise in the PV fields everywhere within the domain of interest. However, we emphasize that this is not necessarily the case in all runs. In the subtropics the difference is maximized between 5° and 18°N. Comparing Figure 4 to Figure 3, it appears that the structure of equivalent length tracks the entrainment rates quite well in the upper levels of the winter hemisphere. In particular, the double-peak vertical structure in the FVDAS entrainment rates is reflected in the subtropical equivalent length. The entrainment rate is relatively small at the same levels where equivalent length is relatively small. This is an indication

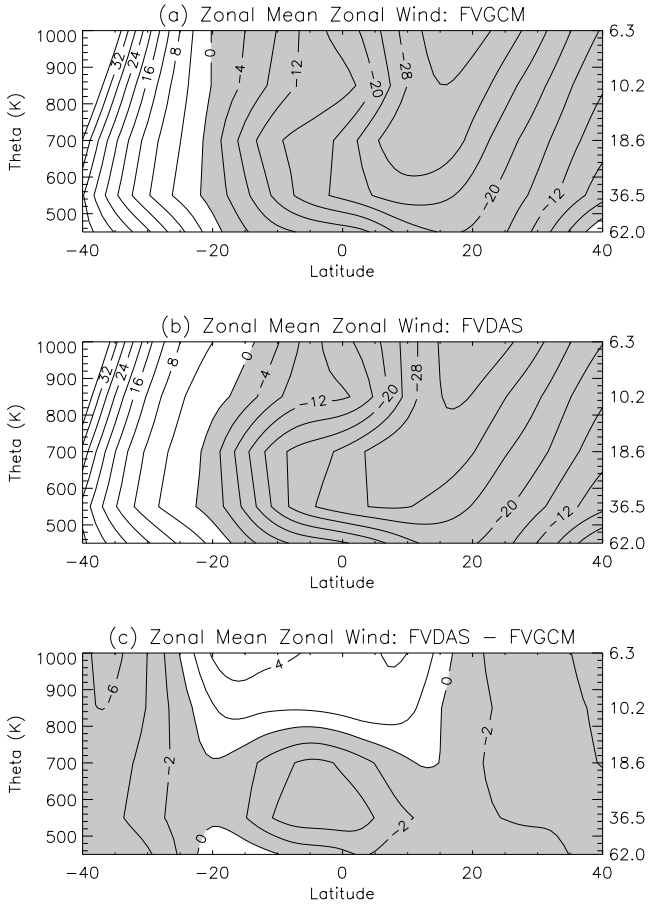


**Figure 5.** Isentropic zonal mean wind in July 1998 from the (a) FVGCM, (b) FVDAS, and (c) difference between FVDAS and FVGCM. Pressure levels (in hPa) corresponding to the isentropes used in this study are shown on the right-hand axis.

that the excess of isentropic subtropical transport and the structure in the equivalent length may have a common source.

### 3.2. July 1998

[20] In July 1998 (Figure 5) in the equatorial region of both the FVGCM and FVDAS a layer of easterly shear appears at the bottom of the domain, while a layer of westerly shear can be seen at around the 750-K isentrope. Weak vertical shear prevails at the top of the domain. In contrast to January 1998 (Figure 1), in which the FVDAS zonal mean wind has an easterly bias compared to the FVGCM zonal mean wind throughout most of the domain, the FVDAS equatorial zonal mean wind in July 1998 has a westerly bias at the upper half of the domain. This is due to a deficit of westerly momentum in the FVGCM since there is no mechanism to generate and maintain a westerly shear in the model. Another noteworthy difference is that the zero-wind line in the FVGCM is almost vertical, while that in the FVDAS slopes toward the equator above 550 K. At 1000 K the zero-wind line is located at 13°S in the FVDAS, compared to 20°S in the FVGCM. Moreover, at 1000 K the magnitude of the easterly wind at the equator is smaller in the FVDAS.

[21] Figure 6 shows entrainment rates in July 1998. Once again, there is almost no entrainment from the summer (northern) hemisphere in the FVGCM, except at 750 K, where it is <0.5%. Entrainment rates from the winter (southern) hemisphere are also insignificant, except at 1000 K, where it is actually larger than that in the FVDAS. An examination of equivalent length showed that the zero-wind line at 1000 K was located at a region with larger equivalent length compared to that in the FVDAS. This indicates that the actual transport barrier in the FVGCM may lie equatorward of the zero-wind line. *Neu et al.* [2003] showed just such a case in July 1998, when the actual transport barrier at 6.8 hPa (close to 1000 K near the equator) was located near the equator while the zero-wind line was located at ~15°S. When we used the equator as the reference latitude, entrainment rates in both the FVGCM and the FVDAS vanished. For the FVDAS, however, the entrainment rates are quite large in both hemispheres. The winter hemisphere rates are close to those in January 1998 except at 550 K, where they are much smaller, and at 1000 K, where they are much larger. Below 700 K and at 850 K the dominant contribution comes from the summer hemisphere. At 700 and 1000 K the contribution from the winter hemisphere dominates. Similar to what happened in January 1998, when a reference latitude closer to the equator was used at 450 K, the entrainment rate decreased, indicating that the actual barrier at 450 K may be closer to the equator than the zero-wind line, which in this case was located at around 17°S. However, regardless of the choice of reference latitude, the FVDAS rates at 450 K remain

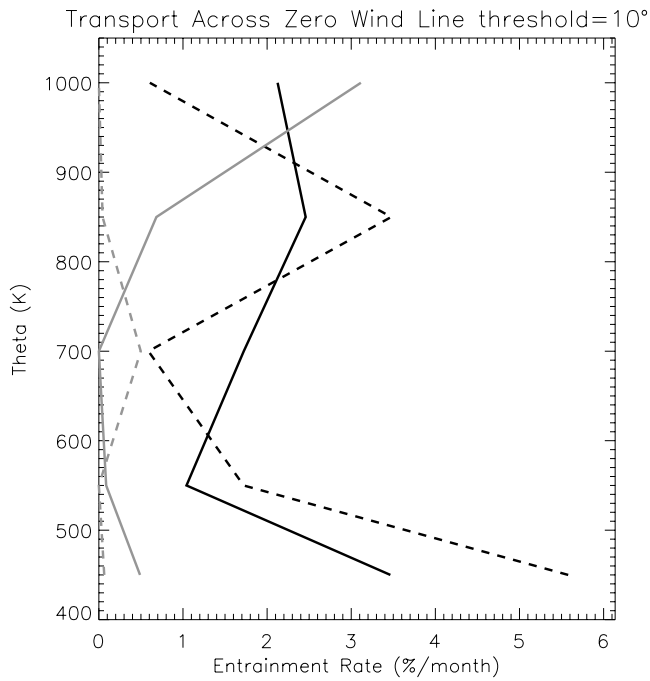


**Figure 6.** Entrainment rates obtained by summing three 10-day calculations in July 1998. Results for the FVDAS (black curves) and results for the FVGCM (grey curves) are shown. The rates from the winter (southern) hemisphere (solid curves) and summer (northern) hemisphere (dashed curves) are shown.

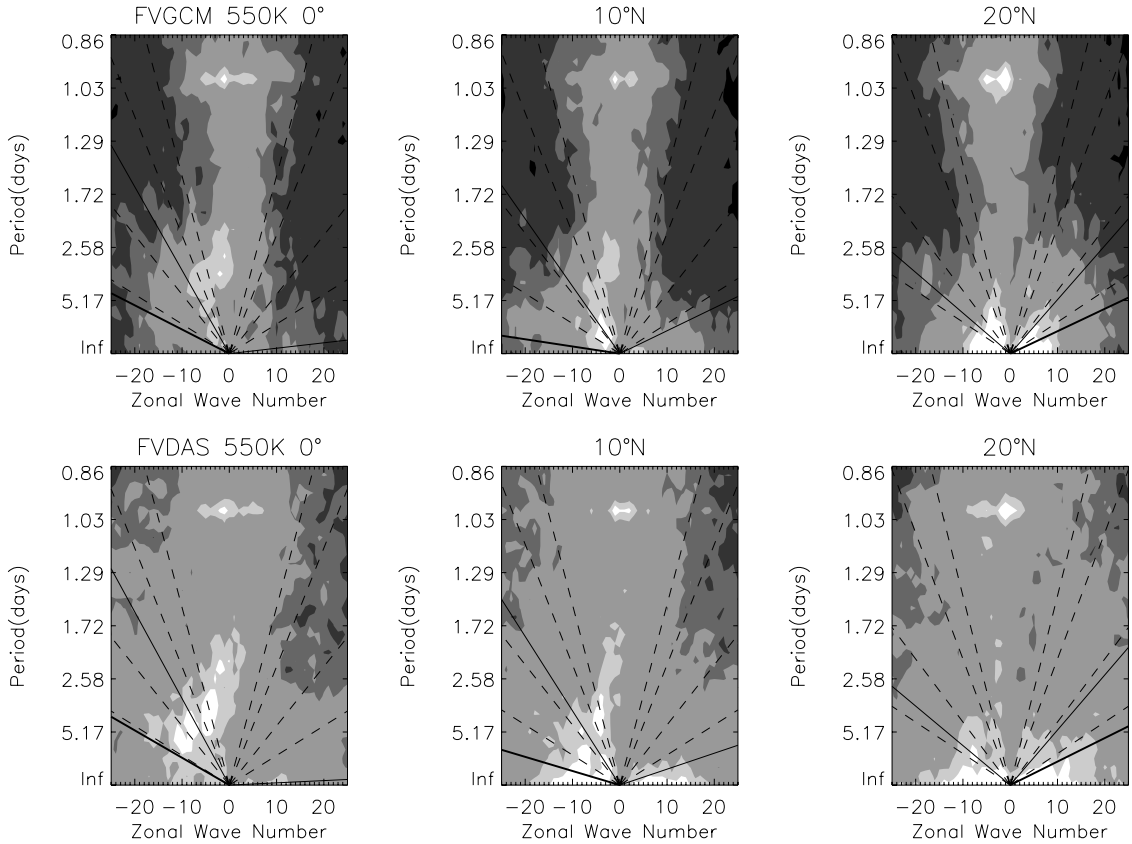


**Figure 7.** Power spectra of the meridional wind for the FVGCM and FVDAS. Shown in each diagram are the −5, −10, −15, −30, 30, 15, 10, and 5 m $s^{-1}$ phase lines (dashed lines). The contour values are 0.0, 0.001, 0.005, 0.01, 0.05, and 0.1 ($m^2$ $s^{-2}$), with the brightest contour having the largest value. The phase line with speed equal to the time mean zonal mean wind, $U$ (thick solid line), and phase lines defined by $C = U - 10$ m $s^{-1}$ and $C = U + 5$ m $s^{-1}$, respectively (thin solid lines), are shown.

]nonvanishing. Another point to note is that in both January 1998 and July 1998, the entrainment rates in the Northern Hemisphere are relatively large. In fact, there is a cell of large equivalent length (not shown) centered at around 850 K in the northern subtropics in July 1998. This coincides with the relatively large entrainment rate. Once again, the bottom line is that entrainment rates in the FVDAS are significantly larger than those in the FVGCM except at 1000 K, where the FVGCM rate is slightly larger. It should be noted again that there is no direct wind observation at this isentrope. As in the case in January 1998, when the thresholds were extended to straddle the zero-wind line, the entrainment rates decreased, but the qualitative difference of the two data sets remained the same.

## 4. Diagnosis

[22] The results of section 3, in conjunction with those presented in *Douglass et al.* [2003] and *Schoeberl et al.* [2003], clearly demonstrate that the assimilation process has a substantial impact, leading to larger transport and mixing. Quantitative measures of the mixing have been presented, specifically showing how entrainment into the tropics increases, and the equivalent lengths calculated from PV fields extend in the assimilation process. In this section, the causes of these differences are sought. We begin by providing a phenomenological understanding through the analysis of the space-time power spectra of the meridional wind field, which reveals an excess of easterly wave power in FVDAS. We then examine the PV fields to get a better idea of the processes that cause the differences in the power spectra. It should be pointed out that several mechanisms are capable of exciting transient features; it is thus reasonable to examine various causes of the superfluous easterly waves in the assimilated fields. We present only results for January 1998.

### 4.1. Space-Time Power Spectra

[23] We examined the space-time power spectra of the meridional wind in the subtropics using the two-dimensional (2-D) fast Fourier transform (FFT) for the 6-hourly data sets. We also selectively carried out more elaborate calculations to separate traveling waves from stationary waves using the spectral partitioning method of *Hayashi* [1977]. The results are similar. The FFT was used because it facilitates the performance of filtering experiments (described in section 4.1). Figure 7 shows examples of the 2-D FFT spectra for the FVGCM and FVDAS, describing the wave number-period structure of waves at 550 K between the equator and 20°N in January (note thatthe negative wave numbers correspond to easterly wave propagation). The plots are smoothed with a [1, 4, 1] running mean in both the horizontal and vertical directions for clarity of presen-

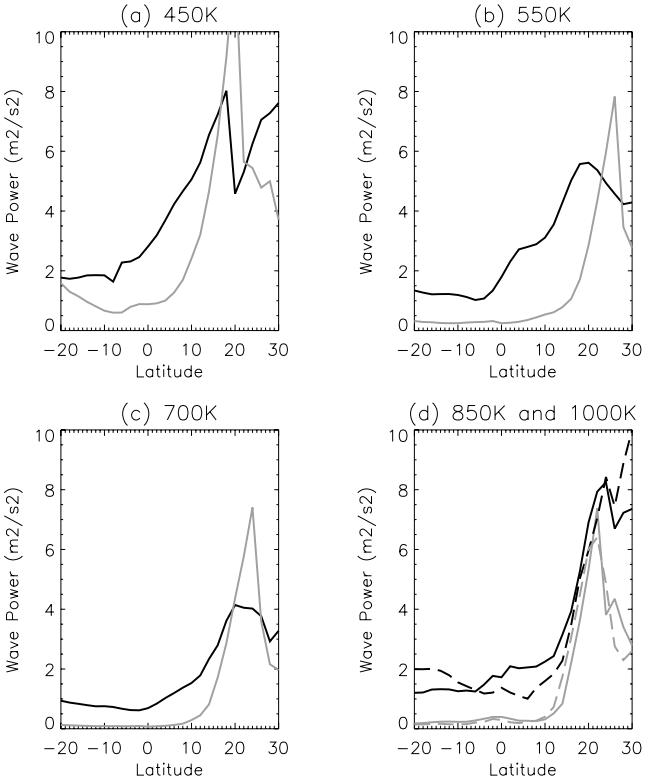


**Figure 8.** Integral spectral power for wave components with phase speeds fulfilling the condition $U - 10$ m s$^{-1}$ < $C < U + 5$ m s$^{-1}$ (area bounded by the two thin solid lines in Figure 7, counterclockwise from the negative half) for theta surfaces of concern. The results for the FVDAS (black curves) and FVGCM (grey curves) are shown. The results at (a) 450, (b) 550, (c) 700, and (d) 850 and 1000 K are plotted. Results for 1000 K (dashed curves) are shown.

tation. In both the FVGCM and FVDAS, easterly traveling waves dominate the spectra southward of 20°N at all levels examined, with the exception of 450 K (not shown), where easterly waves dominate only southward of 10°N. Overall, spectral power in the FVDAS is larger than in the FVGCM for most spectral elements, revealed by the larger areas occupied by contours of equal spectral values in the FVDAS. The locations of some dominant bands in the spectra are similar in the two data sets, e.g., the band of strong easterly traveling waves around the equator at 550 K with rather long wavelengths (wave number ~5) and periods (~3 days). Similar features can be seen at higher isentropes, with some shift toward larger wave numbers and shorter periods in the FVDAS. The difference is usually that the spectral power is significantly stronger in the FVDAS within these bands. Note that in both the FVGCM and FVDAS most of these bands of strong spectral power have phase speeds easterly (clockwise) relative to the zonal mean wind, at least in the tropics. North of ~20°N the spectra reveal a transition to the midlatitude mode (plots not shown), where westerly propagating features dominate. Recall that the zero-wind lines in both data sets are located just south of 20°N (Figure 1). Between 20°N and the equator, the FVDAS exhibits regions of large power not seen in the FVGCM. For instance, at 550 K in the FVDAS (Figure 7), there is a band with large power at 10°N, centered around wave number 8, with a period longer than 5 days. At the equator, there is a pronounced region of similar wave number with a period around 5 days. At 850 K (plots not shown) we see elongated streaks of strong spectral elements running parallel to the mean wind line around the equator. Animation of the PV field shows that this is partially related to the insertion of PV patches which are advected away with the wind as a result of analysis increments at some rawinsonde stations.

[24] Of particular interest is the integral power of the space-time spectra of spectral elements with phase speeds close to the mean zonal velocity. It is this part of the spectra that causes most of the transport [e.g., *Bowman and Hu*, 1997], as it can generate large meridional parcel displacements. Figure 8 shows the integrated spectral power for wave components with phase speeds fulfilling the condition $U - 10$ m s$^{-1}$ < $C < U + 5$ m s$^{-1}$ (where $U$ is the zonal mean wind and $C$ is the phase speed) for the five isentropes. This corresponds to the area bounded by the thin solid lines (counterclockwise from the negative half) in Figure 7. South of 15°N the FVDAS spectral power within this spectral band far exceeds that in the FVGCM at 450 K. For the 550-, 700-, and 850-K theta surfaces this dominance extends further to the north. At the three lower levels the modeled spectra show a distinct peak northward of the zero-wind line, while the assimilated data have substantial power at a broader range of latitudes. At 1000 K the difference of the integral spectral power between the two data sets is small between 15° and 20°N but still quite large equatorward of 15°N. This increase in spectral power coincides with the band of large equivalent length in the subtropics. It is the stirring of these eddy features that lead to the large equivalent length in the subtropics.

[25] To prove that most of the transport rates can be accounted for by this part of the spectra, we conduct a new set of calculations with those waves with phase speeds closely matched to the zonal mean wind (the portion of the spectra bounded by the thin solid lines in Figure 7) filtered out in the region between 10°S and 16°N. The zonal mean wind field is kept intact. The entrainment rates based on the trajectory calculations using this filtered data set show huge reductions compared to the unfiltered case (Figure 3). As shown in Table 1, entrainment rates from the Northern Hemisphere drop substantially. Results using larger threshold values are similar. This confirms that the dominant transport across the subtropical barrier arises from this portion of the wave spectrum, so that the excessive power here in the assimilated data set is the reason for the enhanced meridional transport.

### 4.2. Upward Propagating Features

[26] Having established the cause of the excessive entrainment, it is appropriate to ask how this arises in the assimilation process. For simplicity the first possibility

**Table 1.** Entrainment Rates From the Northern Hemisphere in the FVDAS

| | 450 K | 550 K | 700 K | 850 K | 1000 K |
|---|---|---|---|---|---|
| Original (Figure 3) | 3.7% | 8.9% | 2.3% | 3.6% | 0.61% |
| Filtered | 0.72% | 0.19% | 0.26% | 1.2% | 0.14% |

examined is whether disturbances induced by upward propagation of waves excited by the insertion of data in the troposphere contribute substantially. To investigate this possibility, a data-withholding experiment with stratospheric data withheld was performed with the FVDAS. This experiment omitted all data above 100 hPa in the (otherwise identical) assimilation system. The entrainment rates for this experiment are comparable to those in the FVGCM, except at 750 K. At 450 K the rate is 0.76%, compared to 0.37% in the FVGCM and 3.7% in the FVDAS. At 550 K the corresponding rates are 0.22, 0.05, and 8.9%. At 750 K they are 1.8, 0.58, and 2.3%. Note that the increase of the entrainment rate at 750 K is due solely to a large value during the first 10-day period. Otherwise, the values are comparable to those in the FVGCM. At 800 K the values are 0.17, 1.1, and 3.6%. At 1000 K they are 0.60, 0.64, and 0.61%. The power spectra of the meridional wind are similar to the descriptions given in section 4.1, with the amplitudes between those in the FVGCM and FVDAS. An examination of the integral spectral power shows that on the 550- and 700-K theta surfaces, the integral spectral power in the data-withholding experiment is indeed larger than that in the FVGCM but much smaller than that in the FVDAS south of 15°N. Examinations of the divergence fields in the lower stratosphere show that they resemble the corresponding fields in the full FVDAS, with a lot of streak-like features in the tropics. In contrast, the divergence fields in the FVGCM are much smoother and devoid of these streak-like features. This indicates that upward propagating gravity waves are indeed excited by the data assimilation process. On the basis of these results, while we cannot completely rule out the contribution of upward propagating waves generated in the process of data assimilation to the excess of subtropical transport in the lower stratosphere, we can say with reasonable confidence that their impact is limited.

[27] This interpretation is physically reasonable. Regardless of its source, any upward propagating wave will have only a reversible impact on transport in a region where it is not damped or breaking [e.g., *Andrews and McIntyre*, 1976]. So even though a spectrum of waves is apparently excited by data insertion in the troposphere, the impact of these waves on transport across the subtropical barrier is minimal. This is an important statement, because even though the stratospheric meteorological fields are disturbed by the existence of these waves, they do not substantially impact the irreversible transport.

### 4.3. Comparative Impacts of Data Insertion in the Midlatitudes and the Subtropics

[28] We examined the horizontal component of the Eliassen-Palm (E-P) flux [*Andrews et al.*, 1987] and found that its subtropical magnitude in the northern (winter) hemisphere is larger in the FVGCM compared to the FVDAS (plots not shown). Likewise, south of 30°N the subtropical value of the E-P flux divergence in the FVGCM is comparable to or larger than the FVDAS. This indicates that in the FVDAS waves do not break more vigorously into the subtropics. A second set of data-withholding experiments is now discussed. These runs

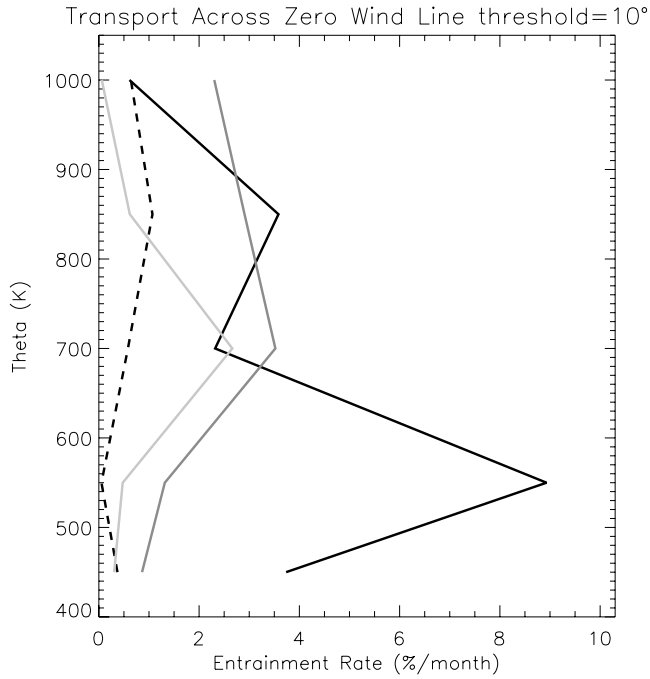


**Figure 9.** Winter (northern) hemisphere entrainment rates for data-withholding experiments. Results from the full FVDAS (black curve), fvdas15 (dark grey curve), fvdas30 (light grey curve), and FVGCM (black dashed curve) are shown.

address the issue of whether the larger wave power in the FVDAS is due to local data insertion or more vigorous wave breaking as a result of data assimilation in the midlatitudes. Two experiments were conducted. The first was done with all data between 30°S and 30°N and above 100 hPa withheld (hereafter fvdas30); that is, no data were used in the "tropical" stratosphere. The second was done using data only between 15°S and 15°N and above 100 hPa (hereafter fvdas15), i.e., data only in the "tropical" stratosphere. Figure 9 shows the entrainment rates of these experiments together with that from the full FVDAS for comparison. Only results in the winter hemisphere are shown. The black curve indicates results from the full FVDAS, the dark grey curve indicates results from fvdas15, and the light grey curve indicates results from fvdas30. Rates in fvdas15 exceed the corresponding rates in fvdas30 at all levels. Note that there is a large reduction at 550 K in fvdas15 compared to the full FVDAS. We believe this may be partially due to the shift of the zero-wind line toward the equator. In the full FVDAS the zero-wind line at 550 K was located at around 16°N, in fvdas15 it was located at around 6°N, in fvdas30 it was located at around 10°N, and in the FVGCM it was located at around 14°N. With the reference latitude set at 6°N, the rate calculated is essentially the rate of cross-equator flow. If more data at higher latitudes are included (e.g., using data between 25°S and 25°N and above 100 hPa only), the entrainment rates increase correspondingly. The results of these experiments are ambiguous owing to the change in the locations of the zonal mean wind, and without further experiments we are not able to draw a definitive conclu-

sion based on them. Nevertheless, the larger rate in fvdas15 compared to that in fvdas30 shows that local data insertion does play a significant role in generating excessive subtropical transport.

### 4.4. Barotropic Instability and PV Patches Related to Data Insertion

[29] Having examined the space-time power spectra and the impacts of data insertion in various regions, we turn to a detailed analysis of the vorticity and PV fields to obtain some clues regarding the impact of locally generated features on subtropical transport. *Sobel et al.* [1997] discuss spurious transport in the "contour-crossing" algorithm from the perspective of error in the PV fields. *Shuckburgh et al.* [2001] show that in the westerly phase of the QBO the subtropical barriers are weakened, as a result of weak barotropic stability or instability formed at the flanks of the equatorial westerly jet. We found that weakened barotropic stability is also present in the subtropics of the FVDAS data set for the period studied. In our case these features are directly related to the analysis increment. Figure 10 shows the 40-hPa (around 550 K) monthly mean meridional gradient of the absolute vorticity fields in the FVGCM and FVDAS, as well as the monthly mean meridional gradient of the vorticity analysis increment for the FVDAS. The 20°N latitude circle is plotted as a line of reference. Comparing the FVGCM field (Figure 10a) and the FVDAS field (Figure 10b), we notice that the former has a smaller variance and smoother features. In contrast, large patches of negative gradient can be seen south of 20°N in the latter. Owing to the zonal orientation of these features, they weaken large-scale stability and in some cases, appear to be barotropically unstable (zonal mean plots not shown). More strikingly, these features have a very strong spatial correlation to the analysis increment shown in Figure 10c. It appears that local data insertion indeed plays a major role in forcing regions of instability in the subtropics, leading to the vast increase in the wave power shown in section 4.1. It also reinforces the result of our data-withholding experiments that local data insertion plays a significant role in the excess of subtropical transport. While it may be argued that the unstable structure driven by the data insertion was real, there are good reasons that they are suspect. First of all, since direct wind observations are sparse in the subtropics, derivatives of the analyzed winds necessarily exhibit more noise compared to GCM winds. Where analysis increments are large, the discrepancies between data-rich regions and data-poor regions are also large. As a result, there is no good reason to believe in the truthfulness of the vorticity increment, which is the gradient of the wind increments. Second, analysis increments are strongly influenced by the choice of forecast error covariance model. The use of different forecast error covariance could lead to very different transport and mixing. We have indeed verified this in two different versions of the FVDAS using different forecast error statistics. Last but not least, some of these features were generated by wind increments inferred from mass data assuming a certain wind-mass error covariance relation. In fact, data-withholding experiments using only mass data showed that subtropical transport and mixing in the FVDAS were also larger compared to the FVGCM. These are good circumstantial evidences to be skeptical of the instability structure generated in the process of data assimilation.

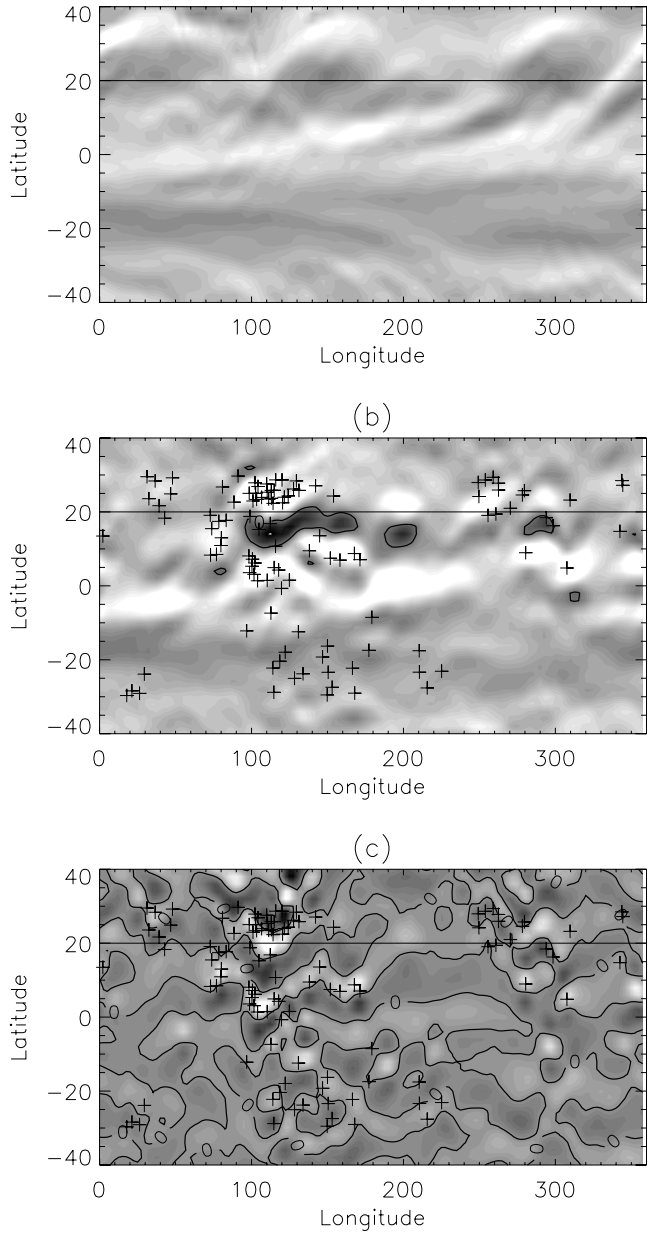


**Figure 10.** Monthly mean meridional gradient of absolute vorticity in the (a) FVGCM and (b) FVDAS at 40 hPa. (c) Monthly mean meridional gradient of the vorticity analysis increments for the FVDAS. More positive values (light shading) and negative values (dark shading enclosed in the solid contours (zero value)) are indicated in Figures 10b and 10c. Since the emphasis is on the correlation between the features in Figures 10b and 10c, we do not plot the values of the color scale. The line across 20°N is plotted as a visual reference.

[30] In addition to the weakening of large-scale stability, due to the lack of PV conservation, data insertions act as sources of PV (and vorticity) with various spatial scales. To demonstrate this, we generate an advected PV field for a particular date using the technique of RDF [e.g., *Sutton et*

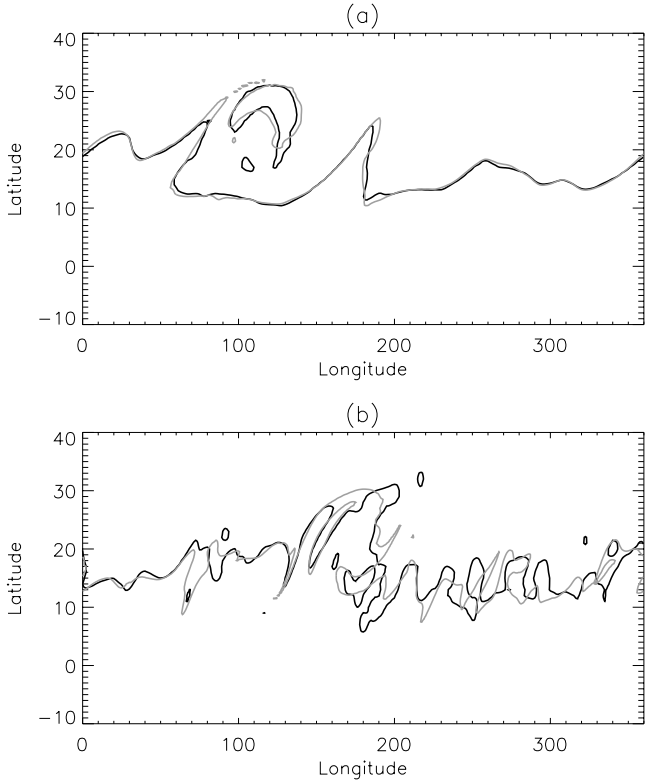


**Figure 11.** (a) The 1-day advected (black curve) and model (grey curve) PV contours with value equal to 40 potential vorticity units at 700 K for the FVGCM at 0000 UT on 31 January 1998 and (b) 1-day advected contours with value equal to 40 potential vorticity units at 700 K for the FVDAS at 0000 UT on 31 January 1998 (black curve) and the analyzed PV contour (grey curve) for the FVDAS.

*al.*, 1994] and compare it to the concurrent analyzed PV field. The advected PV field is generated by initializing a set of parcels on a global grid. These parcels are advected backward in time, and the PV values they encounter along the trajectories are recorded. Plotting the PV values picked up at a particular time along the trajectory back on the day of initialization shows us how the PV field would look if it had been conserved during the advection period. For the FVDAS a comparison of the advected and the analyzed PV fields shows that even for timescales as short as one day, the two fields differ significantly in the tropics, where the planetary vorticity is small. Figure 11a shows the 1-day advected and model PV contours with value equal to 40 potential vorticity units at 700 K for the FVGCM at 0000 UT on 31 January 1998. Both contours roll up in the left half of Figure 11a, in a classical scenario of critical layer wave breaking. The two contours coincide with each other very well except where the secondary vortex detaches from the main contour. This comparison shows the limit of PV conservation for the model dynamics, which is restricted essentially by the finite resolution of the model in the *Lin and Rood* [1996] transport code. In contrast, the scenario for the FVDAS is quite different (Figure 11b). To begin with, compared to the FVGCM, there is much more structure in the FVDAS, the PV contours being replete with short, north-south oriented streaks. Further, the differences between the advected and the analyzed contours are much more pronounced, especially in regions where these short streaks occur, e.g., between 50° and 150°E. These features are intimately related to the analysis increment, as shown in Figure 10. The discrepancies between the analyzed and the advected PV contours result from the process of data insertion, in which PV is not conserved. Animations of the PV field show significant numbers of PV patches being created during the data assimilation process, then being advected around the globe by the zonal mean wind. While the process of data assimilation inserts these PV patches throughout the globe, the problem is more serious in the tropics owing to the diminishing values of the planetary vorticity in the tropics.

[31] It should be pointed out that forcing of the model by the analysis increments alone does not automatically lead to an increase of entrainment rate. It is only in conjunction with wave-breaking processes at higher latitudes that excessive entrainment rates into the tropics are created. Unstable regions around the locations of wave breaking alter particle trajectories through changes in local meridional velocity as well as the emission of spurious waves, increasing the entrainment rates. In other words, it is the cascading effect of wave breaking and the regions of instability forced by analysis increments that cause large net meridional displacements.

## 5. Discussion

[32] Recent studies of subtropical transport using trajectory-related methods as well as CTMs show that assimilated (or analyzed) data sets exhibit significantly larger mixing and entrainment rates compared to GCM data sets [e.g., *Tan*, 2000; *Douglass et al.*, 2003; *Schoeberl et al.*, 2003]. In this study, we examined short-term isentropic transport using Goddard Space Flight Center's (GSFC's) FVDAS and FVGCM to provide a more detailed account of how the process of data assimilation alters the dynamics of the underlying GCM and how this leads to the excess of lower stratospheric mixing and transport in the subtropics.

[33] We show that the excess of subtropical transport is related to the proliferation of eddy features in the subtropics, and we examined various possibilities that may cause this, by withholding data in selected regions. These include the generation of upward propagating features when stratospheric data are withheld, equatorward propagating features when tropical and tropospheric data are withheld, and meridionally confined features when only data in the tropical stratosphere are used. While each of these cases leads to the increase of the subtropical entrainment rate compared to GCM experiments, the impact of data assimilation in the tropical lower stratosphere appears to be the most significant. However, it was not possible to draw a definitive conclusion based on these experiments alone. We observed that the analysis increments of the zonal mean wind exhibit large zonal anomalies, usually with large values in regions where wind observations are available and small values where there are no wind data. This is the consequence of model biases. While any forcing of the GCM with analysis increments of nonnegligible amplitude will disrupt the coherence of the model dynamics, the absence of the QBO in the FVGCM is one of the most

conspicuous biases. Often the analysis increment is not just a small perturbation in the zonal mean wind equation: it can be as large as the physical forcing terms in the Navier-Stokes equations, so that it plays a controlling role. We argue that the episodic insertion of PV (or vorticity) patches in the analysis increment leads to two conditions:

[34] 1. It creates extensive regions with weak or even negative PV gradient, resulting in weak stability or possible instability, between 10° and 20°N, in the assimilated data.

[35] 2. Emission from unstable regions leads to abundance of easterly waves in the tropics. The spectral power for waves with phase speeds close to the zonal mean wind is much larger in the FVDAS than in the FVGCM.

[36] The combination of these two conditions leads to the following results: (1) more frequent meridional incursion of air parcels due to the abundance of waves with phase speeds close to the zonal mean wind and (2) the weakening of the PV restoration mechanism that otherwise restricts the meridional displacement of the parcels. The excess of transport into the tropics from the midlatitudes is the consequence.

[37] In a previous study using an older version of GEOS-DAS [*Tan*, 2000], a similar excess of subtropical transport was found. Moreover, that study found that the subtropical transport rates in the Northern Hemisphere in the UKMO assimilations [*Swinbank and O'Neill*, 1994] for the time period used were also excessively large. On the other hand, *Rogers et al.* [1999] found that while results of aerosol simulations using a three-dimensional transport model (SLIMCAT) driven by the ECMWF and UKMO analyses all suffered from too much transport of midlatitude air into the tropics, the results using ECMWF analyses were closer to observations. All these results show that the problem is not limited to a particular assimilation system, but the degrees of discrepancy are different. As long as significant model bias remains, it is likely that any assimilated data sets will encounter similar problems. A detailed examination of various assimilation systems and results from their parent GCMs will be beneficial to further our understanding of the problem.

[38] The generation of spurious waves in the process of data assimilation has long been recognized as an important issue, and various techniques of initialization have been used to deal with the problem [e.g., *Daley*, 1991]. While an initialization step is not included in the FVDAS or in the current operational version of the GEOS-DAS, earlier versions used the incremental analysis update (IAU) technique [*Bloom et al.*, 1996] to reduce analysis shock. While these versions were also plagued with the problem of excessive subtropical transport and mixing [*Tan*, 2000], the technique has never been implemented in the new version, and the impact is not known. In addition, *Fox-Rabinovitz et al.* [1998] (hereafter FR98) showed that using a diabatic initialization scheme to balance the incremental fields, before using them to force the GCM, is very effective in removing small-scale features in the tropics. From the differences in the PV fields (Figure 9 in FR98) it is clear that the diabatic initialization process has a strong impact on features with zonal wave numbers higher than $\sim$10. Features with scales close to zonal wave number 10 form a significant part of the regions with weak or negative PV gradient, as well as other PV patches due to the analysis increment in the FVDAS data set used in this study. Therefore it is likely that diabatic initialization would reduce regions with weakened stability and thus partially alleviate the problem of excessive subtropical transport. In this regard, it is equivalent to using a filter that is dynamically consistent with the GCM to control the noise.

[39] Another important question is the problem of wind-mass balance in the forecast error covariance model of a data assimilation system [e.g., *Daley*, 1991]. While the assumption of a certain relationship (e.g., geostrophy) between wind and mass errors increases the value of each individual observation, analyses derived from an inappropriate balance relationship have the potential to upset the dynamic of the model, leading to the emission of spurious waves. While the use of an error relationship based on low-order balance of the wind and mass fields may be warranted at higher latitudes, it is questionable in the tropics. In fact, comparisons of analysis increments of the wind fields in regions with and without direct wind observations show that they are often quite different. Since wind observations in the tropics are very sparse, analysis increments of the wind fields at most of the tropical grid points come from temperature observations. Although increments from direct wind observations normally have larger forcing amplitudes, increments coming from temperature observations are not negligible. In fact, as shown in Figure 10, analysis increments derived from the particular implementation of the wind-mass balance in this study forced extensive regions of alternate negative and positive vorticity gradient and thus strongly modified the stability of the background state. The use of an empirical wind-mass balance and the study of its impact on subtropical transport and mixing in the lower stratosphere may yield new insight on the problem.

[40] Important as the tropics are in our efforts to understand the atmosphere, our ability to model them is unsatisfactory. For instance, few models are able to reproduce one of the most prominent phenomena in the equatorial lower stratosphere, the QBO. And even if there is a QBO in the model, there is no guarantee that its profiles will match those in the real atmosphere. This inadequacy, together with the lack of direct wind observations, make the analysis and assimilation of the tropical lower stratosphere a particularly challenging task. The main purpose of data assimilation is to provide global analysis for temperature, geopotential height, and wind fields. The lack of conservation of properties that control the dynamics, epitomized by the anomalous insertion of PV patches, not only causes problems in isentropic mixing, it also creates significant problems in vertical transport [*Gettelman and Sobel*, 2000]. There is no built-in mechanism to ensure a correct modeling of transport characteristics. Short of radically changing the GCM used in the assimilation, we might be able to improve various aspects of the technique of assimilation and analysis to derive a better representation of transport properties. These may include imposing physical constraints to reduce unstable regions created during the assimilation process and adopting a shock reduction procedure (e.g., the IAU) or an initialization process (e.g., as in FR98) to reduce noise in the assimilation fields. However, the ultimate solution of the problem requires a fundamental improvement of our modeling ability and the availability of direct wind observations in the tropics and how they are assimilated. It is only when such problems are solved that dynamical fields from data

assimilation can fully realize their promise of simulating the contribution of long-term variations in dynamics to variations in atmospheric composition (e.g., stratospheric ozone simulations).

[41] **Acknowledgments.** We thank Mark R. Schoeberl for providing the trajectory codes used in this study. We also thank two anonymous reviewers whose comments led to significant improvement of the paper. This work was based on the Ph.D. dissertation of *Tan* [2000] done at the State University of New York at Stony Brook.

## References

Allen, D. R., and N. Nakamura (2001), A seasonal climatology of effective diffusivity in the stratosphere, *J. Geophys. Res.*, *106*, 7917–7935.

Andrews, D. G., and M. E. McIntyre (1976), Planetary waves in horizontal and vertical shear: The generalized Eliassen-Palm relation and the mean zonal acceleration, *J. Atmos. Sci.*, *33*, 2031–2048.

Andrews, D. G., J. R. Holton, and C. B. Leovy (1987), *Middle Atmosphere Dynamics*, 498 pp., Academic, San Diego, Calif.

Bloom, S. C., L. L. Takacs, A. M. da Silva, and D. Ledvina (1996), Data assimilation using incremental analysis updates, *Mon. Weather Rev.*, *124*, 1256–1271.

Bowman, K. P., and Y. Hu (1997), Tropical mixing barriers in the lower stratosphere in the Geophysical Fluid Dynamics Laboratory SKYHI model, *J. Geophys. Res.*, *102*, 21,465–21,478.

Chen, P. (1996), The influences of zonal flow on wave breaking and tropical-extratropical interaction in the lower stratosphere, *J. Atmos. Sci.*, *53*, 2379–2392.

Cohn, S. E., A. da Silva, J. Guo, M. Sienkiwiecz, and D. Lamich (1998), Assessing the effects of data selection with the DAO Physical-Space Statistical Analysis System, *Mon. Weather Rev.*, *126*, 2913–2926.

Daley, R. (1991), *Atmospheric Data Analysis*, 457 pp., Cambridge Univ. Press, New York.

Douglass, A. R., M. R. Schoeberl, R. B. Rood, and S. Pawson (2003), Evaluation of transport in the lower tropical stratosphere in a global chemistry and transport model, *J. Geophys. Res.*, *108*(D9), 4259, doi:10.1029/2002JD002696.

Fox-Rabinovitz, M. S., R. B. Rood, D. J. Lamich, R. C. Govindaraju, L. Coy, and C. Weaver (1998), The impact of diabatic initialization on stratospheric analyses, forecasts, and transport experiments, *Q. J. R. Meteorol. Soc.*, *124*, 297–315.

Gettelman, A., and A. H. Sobel (2000), Direct diagnoses of stratosphere-troposphere exchange, *J. Atmos. Sci.*, *57*, 3–16.

Grant, W. B., E. V. Browell, C. S. Long, L. L. Stowe, R. G. Grainger, and A. Lambert (1996), Use of volcanic aerosols to study the tropical stratospheric reservoir, *J. Geophys. Res.*, *101*, 3973–3988.

Gray, L. J., and J. M. Russell III (1999), Interannual variability of trace gases in the subtropical winter stratosphere, *J. Atmos. Sci.*, *56*, 977–993.

Hamilton, K., and L. Yuan (1992), Experiments on tropical stratospheric mean-wind variations in a spectral general circulation model, *J. Atmos. Sci.*, *49*, 2464–2483.

Hayashi, Y. (1977), On the coherence between progressive and retrogressive waves and a partition of space-time power spectra into standing and traveling parts, *J. Clim. Appl. Meteorol.*, *16*, 368–373.

Haynes, P., and E. Shuckburgh (2000a), Effective diffusivity as a diagnostic of atmospheric transport: 1. Stratosphere, *J. Geophys. Res.*, *105*, 2777–2794.

Haynes, P., and E. Shuckburgh (2000b), Effective diffusivity as a diagnostic of atmospheric transport: 2. Troposphere and lower stratosphere, *J. Geophys. Res.*, *105*, 2795–2810.

Joiner, J., and L. Rokke (2000), Variational cloud clearing with TOVS data, *Q. J. R. Meteorol. Soc.*, *126*, 725–748.

Juckes, M. N. (1989), A shallow-water model of the winter stratosphere, *J. Atmos. Sci.*, *46*, 2934–2955.

Juckes, M. N., and M. E. McIntyre (1987), A high resolution, one layer model of breaking planetary waves in the stratosphere, *Nature*, *328*, 590–596.

Kiehl, J. T., J. J. Hack, G. B. Bonan, B. A. Boville, D. L. Williamson, and P. J. Rasch (1998), The National Center for Atmospheric Research Community Climate Model: CCM3, *J. Clim.*, *11*, 1131–1149.

Lin, S.-J. (1997), A finite-volume integration method for computing pressure gradient force in general vertical coordinates, *Q. J. R. Meteorol. Soc.*, *123*, 1749–1762.

Lin, S.-J., and R. B. Rood (1996), Multidimensional flux-form semi-Lagrangian transport schemes, *Mon. Weather Rev.*, *124*, 2046–2070.

Lin, S.-J., and R. B. Rood (1997), An explicit flux-form semi-Lagrangian shallow-water model on the sphere, *Q. J. R. Meteorol. Soc.*, *124*, 2477–2498.

Ma, J., D. W. Waugh, A. R. Douglass, S. R. Kawa, and S.-J. Lin (2003), Evaluation of the transport in the Goddard Space Flight Center three-dimensional chemical transport model using the equivalent length diagnostic, *J. Geophys. Res.*, *108*(D6), 4201, doi:10.1029/2002JD002268.

Minschwaner, K., A. E. Dessler, J. W. Elkins, C. M. Volk, D. W. Fahey, M. Loewenstein, J. R. Podolske, A. E. Roche, and K. R. Chan (1996), Bulk properties of isentropic mixing into the tropics in the lower stratosphere, *J. Geophys. Res.*, *101*, 9433–9439.

Molod, A., H. M. Helfand, and L. L. Takacs (1996), The climatology of parameterized physical processes in the GEOS-1 GCM and their impact on the GEOS-1 data assimilation system, *J. Clim.*, *9*, 1531–1545.

Mote, P. W., K. H. Rosenlof, M. E. McIntyre, E. S. Carr, J. C. Gille, J. R. Holton, J. S. Kinnersley, H. C. Pumphrey, J. M. Russell III, and J. W. Waters (1996), An atmospheric tape recorder: The imprint of tropical troposphere temperatures on stratospheric water vapor, *J. Geophys. Res.*, *101*, 3989–4006.

Mote, P. W., T. J. Dunkerton, M. E. McIntyre, E. A. Ray, P. H. Haynes, and J. M. Russell III (1998), Vertical velocity, vertical diffusion, and dilution by midlatitude air in the tropical lower stratosphere, *J. Geophys. Res.*, *103*, 8651–8666.

Nakamura, N. (1995), Modified Lagrangian-mean diagnostics of the stratospheric polar vortices, Part I, Formulation and analysis of GFDL SKYHI GCM, *J. Atmos. Sci.*, *52*, 2096–2108.

Nakamura, N. (1996), Two-dimensional mixing, edge formation, and permeability diagnosed in an area coordinate, *J. Atmos. Sci.*, *53*, 1524–1537.

Nakamura, N., and J. Ma (1997), Modified Lagrangian-mean diagnostics of the stratospheric polar vortices, 2, Nitrous oxide and seasonal barrier migration in the cryogenic limb array etalon spectrometer and SKYHI general circulation model, *J. Geophys. Res.*, *102*, 25,721–25,735.

Neu, J. L., L. C. Sparling, and R. A. Plumb (2003), Variability of the subtropical "edges" in the stratosphere, *J. Geophys. Res.*, *108*(D15), 4482, doi:10.1029/2002JD002706.

Polvani, L. M., D. W. Waugh, and R. A. Plumb (1995), On the subtropical edge of stratospheric surf zone, *J. Atmos. Sci.*, *52*, 1288–1309.

Rogers, H. L., W. A. Norton, A. Lambert, and R. G. Grainger (1999), Isentropic, diabatic, and sedimentary transport of Mount Pinatubo aerosol, *J. Geophys Res.*, *104*, 4051–4063.

Schoeberl, M. R., and D. L. Hartmann (1991), The dynamics of the stratospheric polar vortex and its relation to springtime ozone depletions, *Science*, *251*, 46–52.

Schoeberl, M. R., and L. C. Sparling (1995), Trajectory modeling, in *Diagnostic Tools in Atmospheric Physics*, *Proc. Int. Sch. Phys. Enrico Fermi*, *124*, 289–305.

Schoeberl, M. R., A. R. Douglass, Z. Zhu, and S. Pawson (2003), A comparison of the lower stratospheric age-spectra derived from a general circulation model and two data assimilation systems, *J. Geophys. Res.*, *108*(D3), 4113, doi:10.1029/2002JD002652.

Shuckburgh, E., W. Norton, A. Iwi, and P. Haynes (2001), Influence of the quasi-biennial oscillation on isentropic transport and mixing in the tropics and subtropics, *J. Geophys. Res.*, *106*, 14,327–14,337.

Sobel, A. H., R. A. Plumb, and D. W. Waugh (1997), Methods of calculating transport across the polar vortex edge, *J. Atmos. Sci.*, *54*, 2241–2260.

Sutton, T. T., H. Maclean, R. Swinbank, A. O'Neill, and F. W. Taylor (1994), High-resolution stratospheric tracer fields estimated from satellite observations using Lagrangian trajectory calculations, *J. Atmos. Sci.*, *51*, 2995–3005.

Swinbank, R., and A. O'Neill (1994), A stratosphere-troposphere data assimilation system, *Mon. Weather Rev.*, *122*, 686–702.

Tan, W. (2000), Representation of lower stratospheric subtropical transport and mixing in data assimilation systems, Ph.D. dissertation, State Univ. of N. Y., Stony Brook, N. Y.

Trepte, C. R., and M. H. Hitchman (1992), Tropical stratospheric circulation deduced from satellite aerosol data, *Nature*, *355*, 626–628.

Waugh, D. W. (1996), Seasonal variation of isentropic transport out of the tropical stratosphere, *J. Geophys. Res.*, *101*, 4007–4023.

A. da Silva, S. Pawson and W. W. Tan, Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Code 900.3, Greenbelt, MD 20771, USA. (asilva@gmao.gsfc.nasa.gov; spawson@gmao.gsfc.nasa.gov; wtan@gmao.gsfc.nasa.gov)

M. A. Geller, Institute of Terrestrial and Planetary Atmospheres, State University of New York at Stony Brook, Stony Brook, NY 11794-5000, USA. (marvin.geller@sunysb.edu)